



CHAT-PUB FEEDBACK

Helper of welfare policy for youth

김강산
양승빈
박진호
전창민



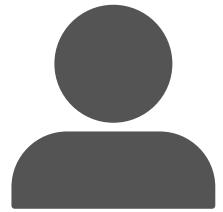


QUESTIONS & ANSWERS

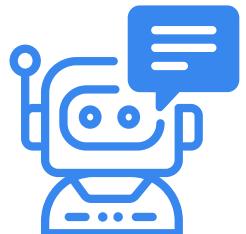
- EXAMPLES OF INPUT AND OUTPUT
- PERSONAL INFORMATION CONSENT
- DETAILS ON DATA CRAWLING
- THOROUGH EXPLANATION ON MODEL

INPUT AND OUTPUT EXAMPLE

USER INPUT

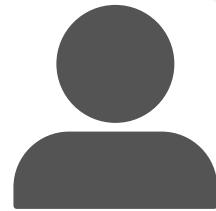


**HI, I'M A 25 AND I CURRENTLY LIVE IN GYEONGGI-DO.
I WONDER IF THERE'S A HOUSING POLICY I CAN BENEFIT FROM?**

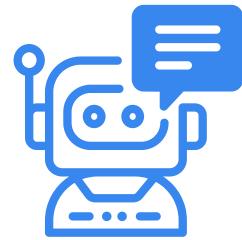


INPUT AND OUTPUT EXAMPLE

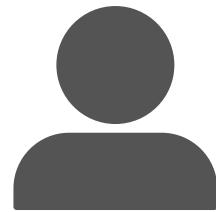
CHAT-BOT OUTPUT



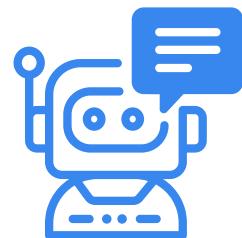
**HI, I'M A 25 AND I CURRENTLY LIVE IN GYEONGGI-DO.
I WONDER IF THERE'S A HOUSING POLICY I CAN BENEFIT FROM?**



COULD YOU TELL ME THE EXACT PLACE OF RESIDENCE?



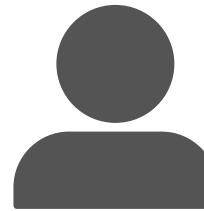
I LIVE IN SOWON



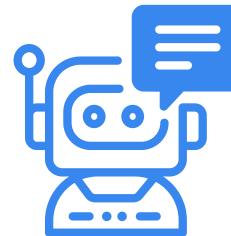
• • • • • • •

INPUT AND OUTPUT EXAMPLE

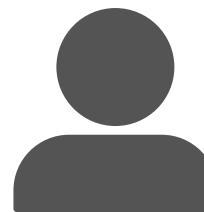
CHAT-BOT OUTPUT



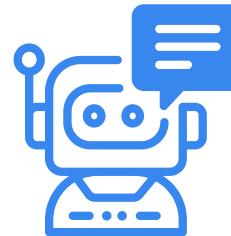
HI, I'M A 25 AND I CURRENTLY LIVE IN GYEONGGI-DO.
I WONDER IF THERE'S A HOUSING POLICY I CAN BENEFIT FROM?



COULD YOU TELL ME THE EXACT PLACE OF RESIDENCE?



I LIVE IN SOWON



HERE IS POLICY THAT MIGHT BENEFIT YOU!

THE POLICY IS CALLED RECRUITMENT OF YOUTH ZONES THAT IS FOR TO PROMOTE HOUSING STABILITY OF YOUNG PEOPLE IN SUWON CITY TO 50% LEVEL OF YOUNG PEOPLE IN SUWON CITY.

YOU HAVE TO BE AT LEAST 19 AND NO OLDER THAN 34. THE RECRUITMENT SCHEDULE FOR THE SECOND HALF OF THE YEAR HAS NOT YET BEEN RELEASED.

PRIVACY PROBLEM

PERSONAL INFORMATION AGREEMENT

Personal Information Consent Form

[Essential] Consent to collect personal information for the use of chat-pub

[Personal Information Agreement]

Collected personal Information	Name、Sex、Student record、Grade、Date of Birth、Contact NO、Address (Korean、Home country)、Family particulars、Academic background、Language ability etc
Purpose of collecting personal information	The information provided is used to recommend the policy you need. 1) Key services: Youth policy selection, recommendation 2) Government agencies: various statistical reports, internships 3) Other duties deemed necessary
Period of retention	Sungkyunkwan University Capstone Team/10years
<Disadvantage in case of refusal> If you refuse agreement to collect and use private information, we may not be able to accept your application for service	

I agree

I disagree

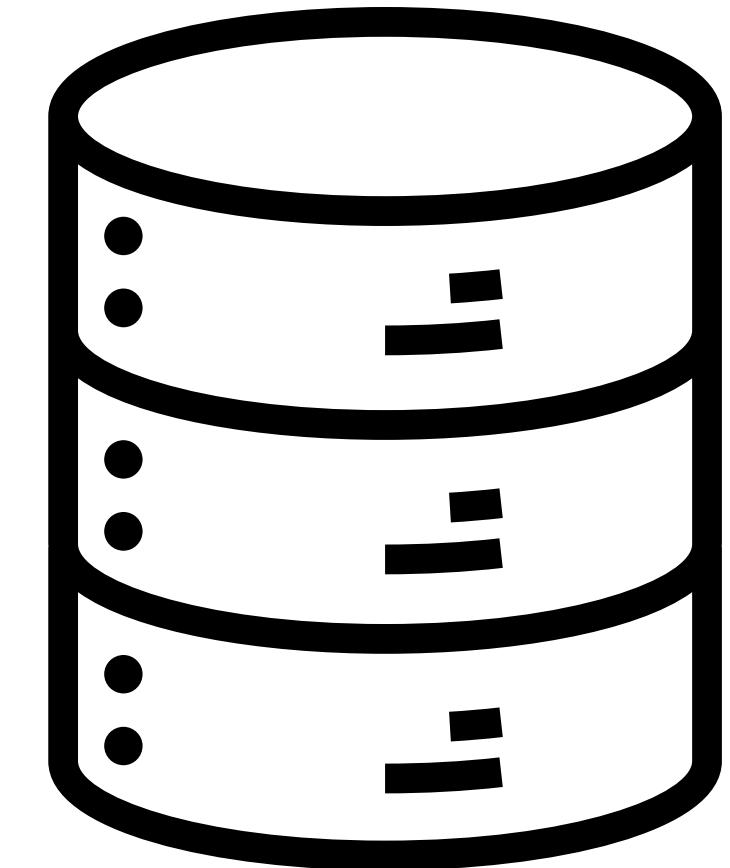
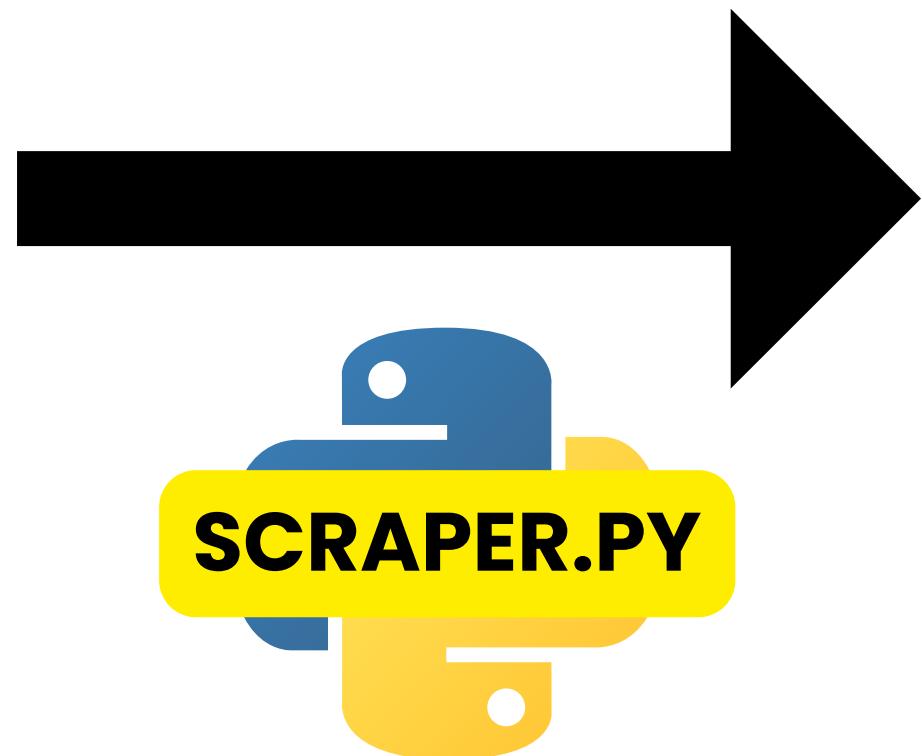
TECHINAL BACKGROUND

DATA CRAWLING & EMBEDDING



- **Raw Data**

- we will fetch data from
<https://www.youthcenter.go.kr/main.do>
by our own scraper, which crawl data from web.

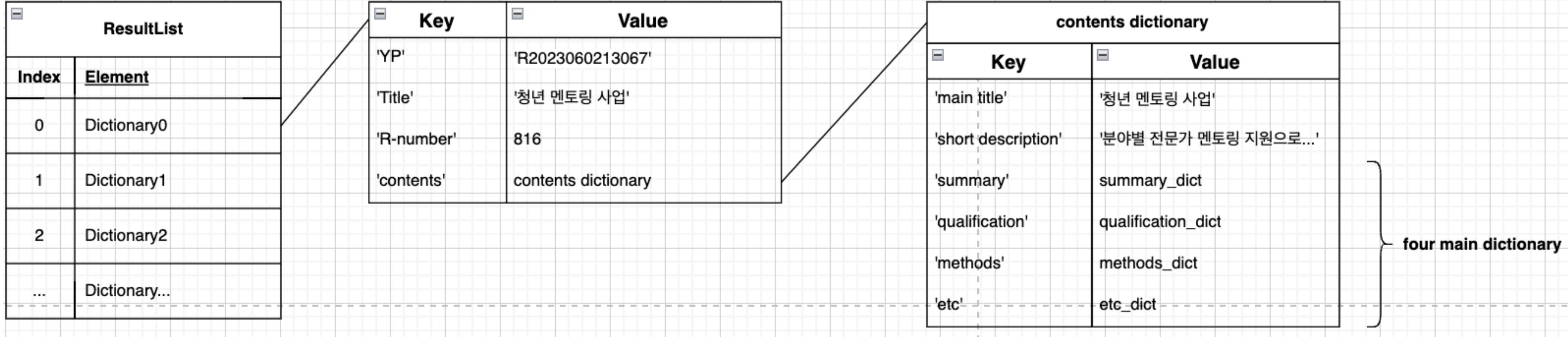


- **Our Data**

TECHINAL BACKGROUND

DATA CRAWLING & EMBEDDING

Data Structure of ChatPub



TECHINAL BACKGROUND

DATA CRAWLING & EMBEDDING

Data Field of four main dictionary

1. summary dict

```
{  
  '정책 번호'  
  '정책 분야'  
  '지원 내용'  
  '사업 운영 기간'  
  '사업 신청 기간'  
  '지원 규모(명)'  
  '비고'  
}
```

2. qualification dict

```
{  
  '연령'  
  '거주지 및 소득'  
  '학력'  
  '전공'  
  '취업 상태'  
  '특화 분야'  
  '추가 단서 사항'  
  '참여 제한 대상'  
}
```

3. methods dict

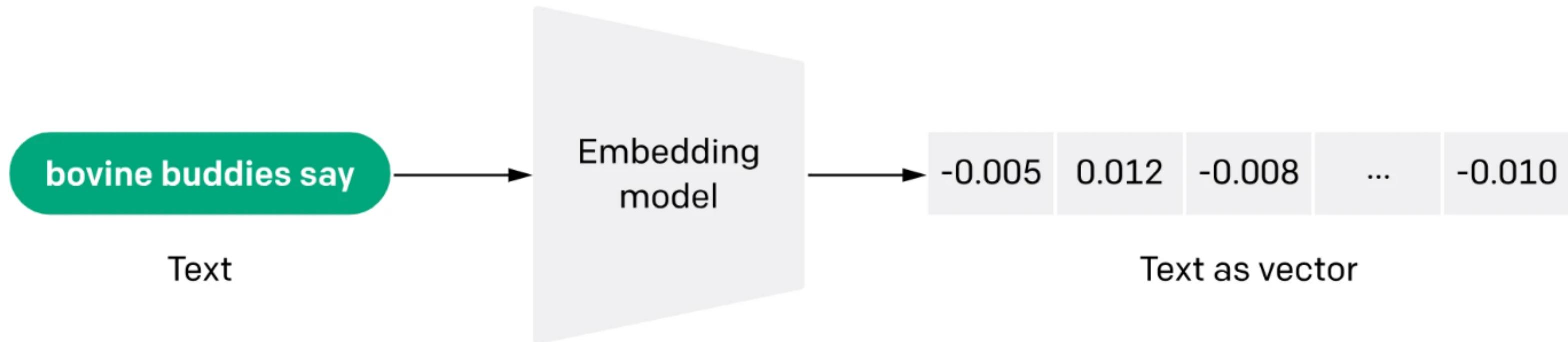
```
{  
  '신청 절차'  
  '심사 및 발표'  
  '신청 사이트'  
  '제출 서류'  
}
```

4. etc dict

```
{  
  '기타 유익 정보'  
  '주관 기관'  
  '운영 기관'  
  '사업관련 참고 사이트 1'  
  '사업관련 참고 사이트 2'  
  '첨부파일'  
}
```

TECHINAL BACKGROUND

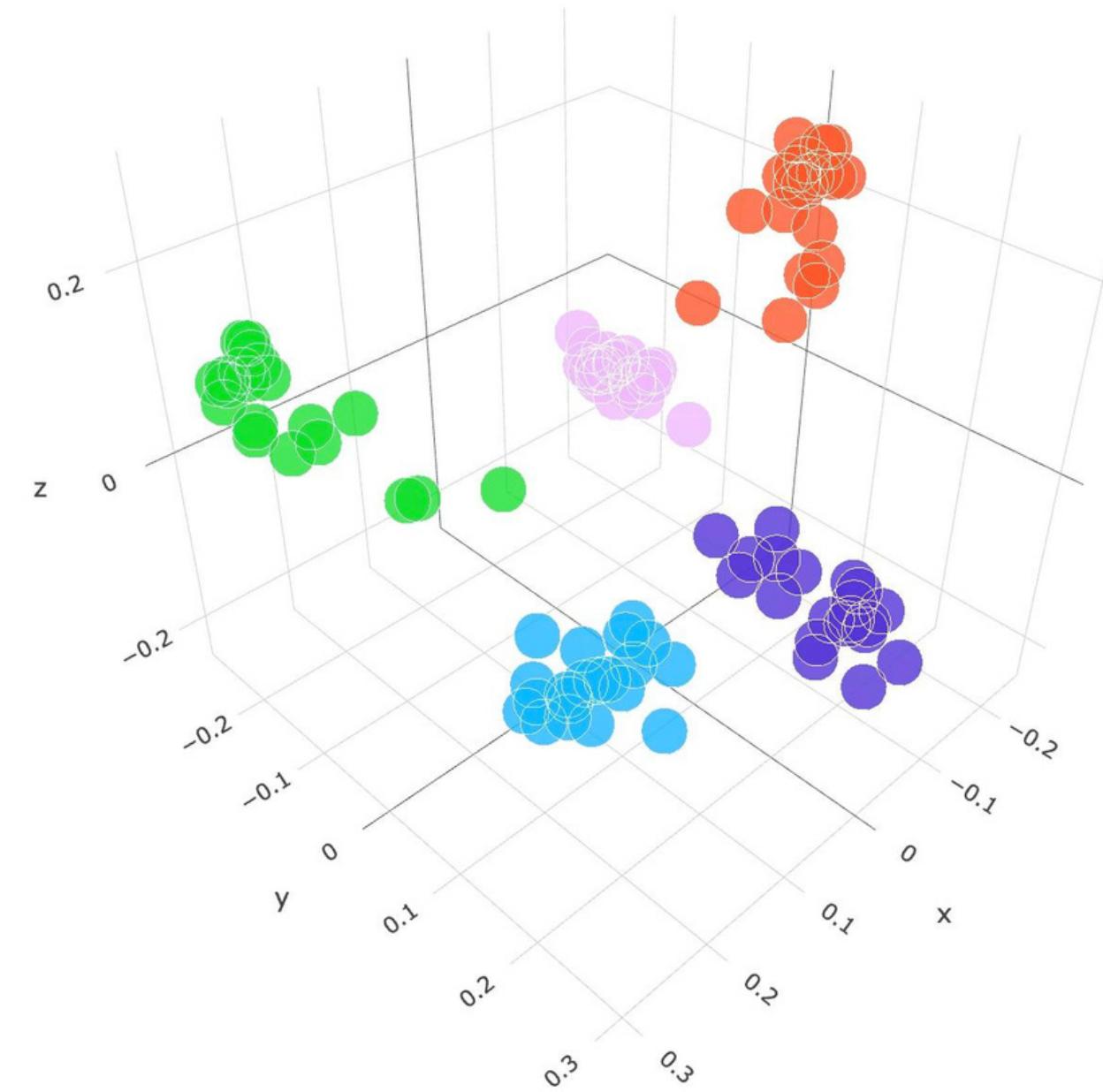
TEXT EMBEDDING



TECHINAL BACKGROUND

TEXT EMBEDDING

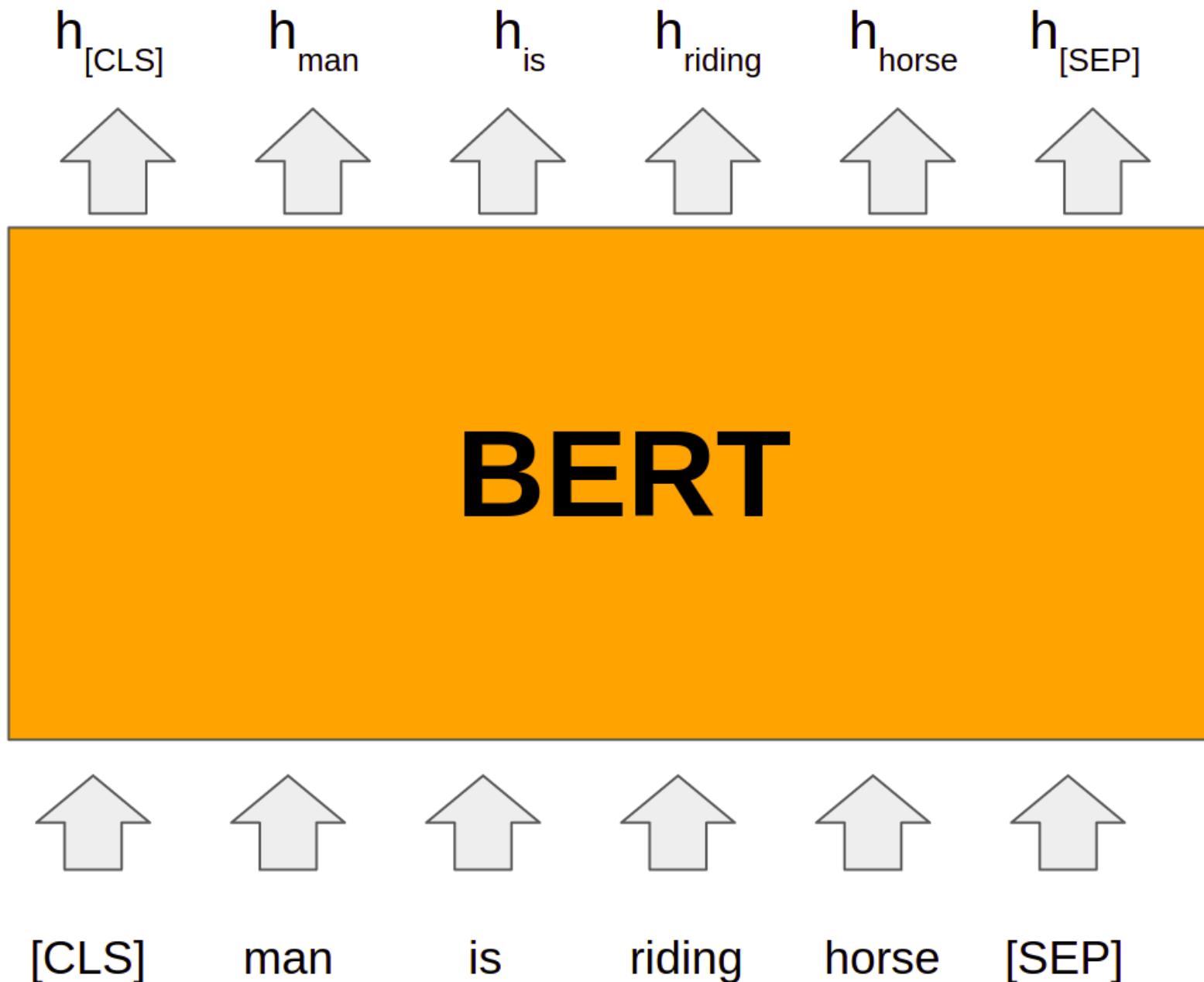
animal athlete film transportation village



- Texts with similar meanings are embedded as vectors nearby.
- Texts with different meanings are embedded as vectors far away.
- We can use BERT-based models to perform text embedding.

TECHINAL BACKGROUND

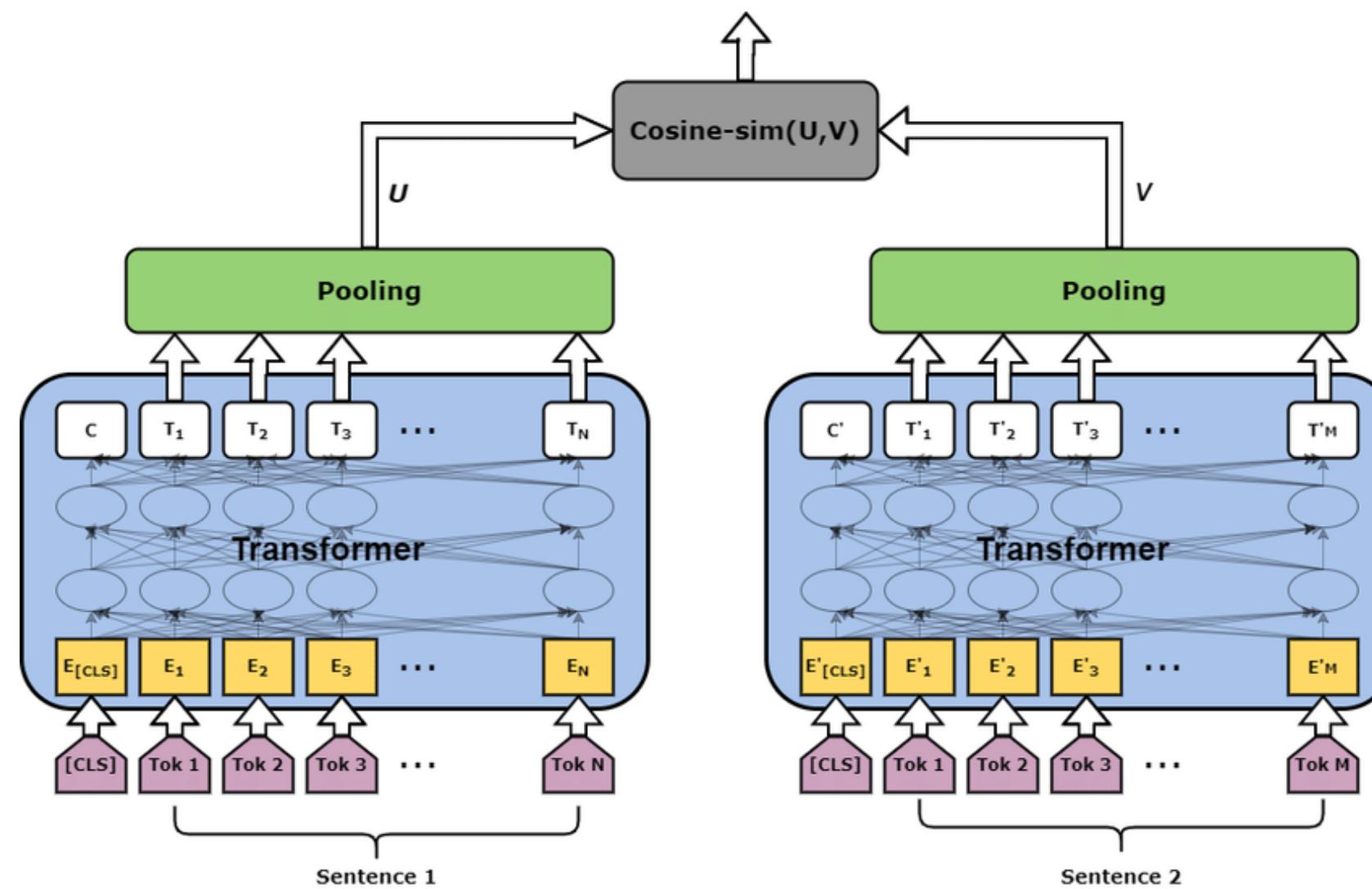
BERT



- BERT is a model that adopts the Encoder structure of the Transformer.
- Create embeddings for text utilizing contextual information

TECHNICAL BACKGROUND

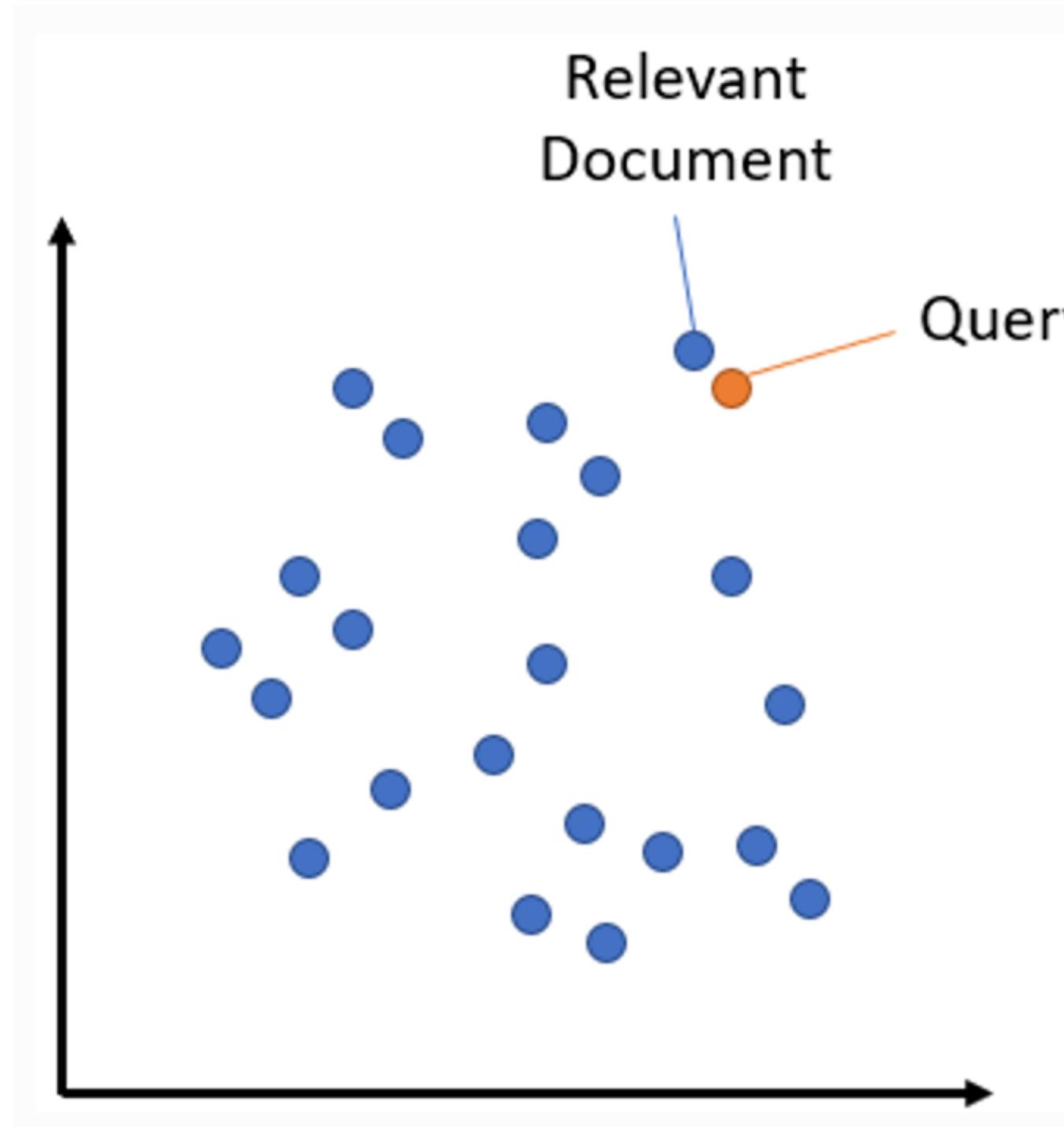
SENTENCE TRANSFORMER



- SentenceTransformers is a Python framework for state-of-the-art sentence, text, and image embeddings.
- After text embedding is performed using the Sentence Transformers (BERT structure), similarity can be measured by cosine similarity.

TECHINAL BACKGROUND

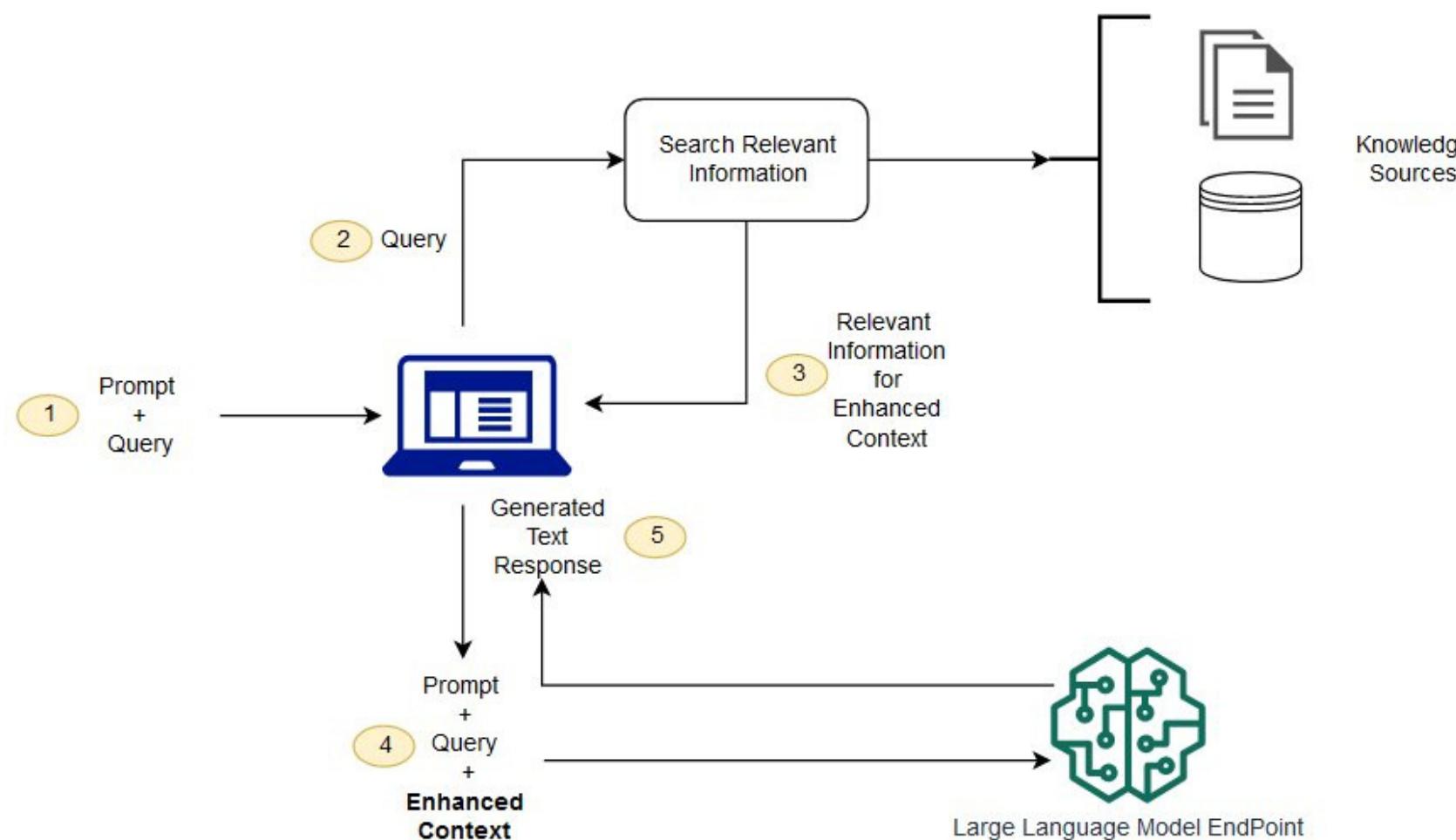
EMBEDDING-BASED RETRIEVAL



- When a user's question is asked, the embedding vector for the question and the embedding vector for the related document are located in a close space.
- Therefore, we can retrieve the document content needed to answer the question using embedding!

TECHINAL BACKGROUND

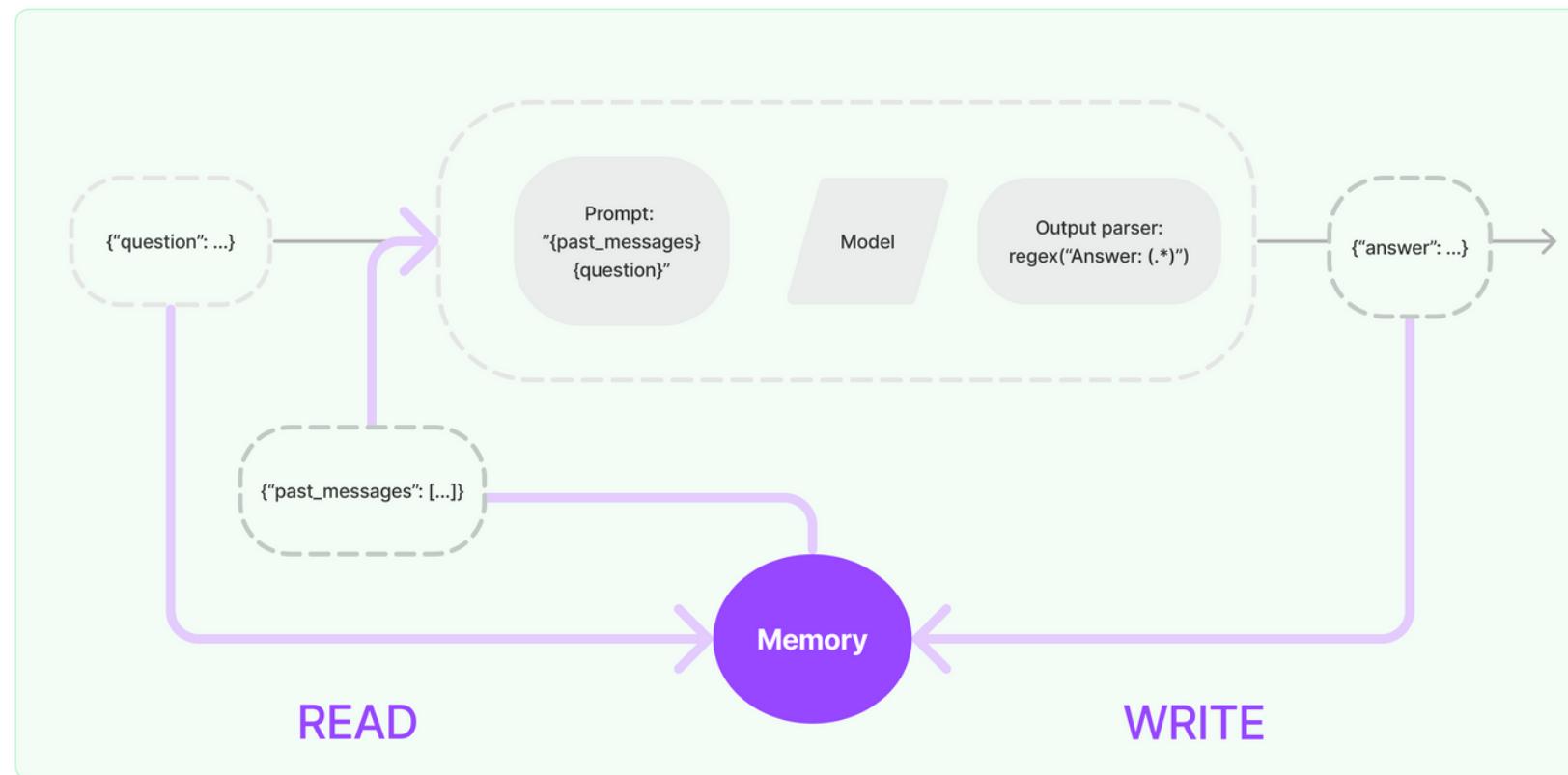
RETRIEVAL AUGMENTED GENERATION



- Retrieval Augmented Generation (RAG) technology enables generative models to use external knowledge (not utilized for training) when performing answers.
- Perform a retrieval for a given user question to extract information from the knowledge store and combine it with a prompt to generate an answer.

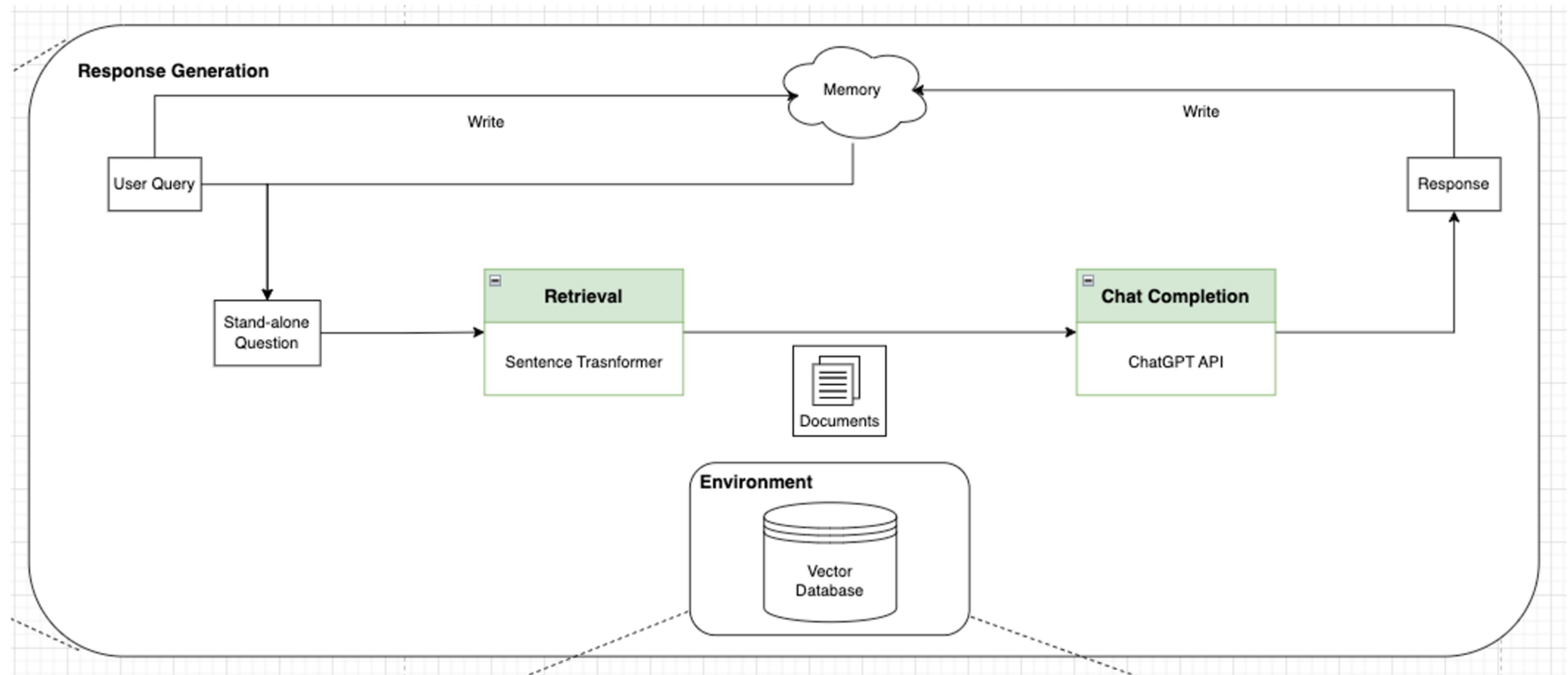
TECHINAL BACKGROUND

MEMORY



- The chatbot should remember the context of the previous conversation and refer to it in generating answers.
- We will utilize the memory capabilities provided by Langchain.
- Performs saving the user's speech and the chatbot's answers. The saved conversation context is utilized to generate Retrieval and answers.ODO

MODEL PIPELINE



MODEL PIPELINE

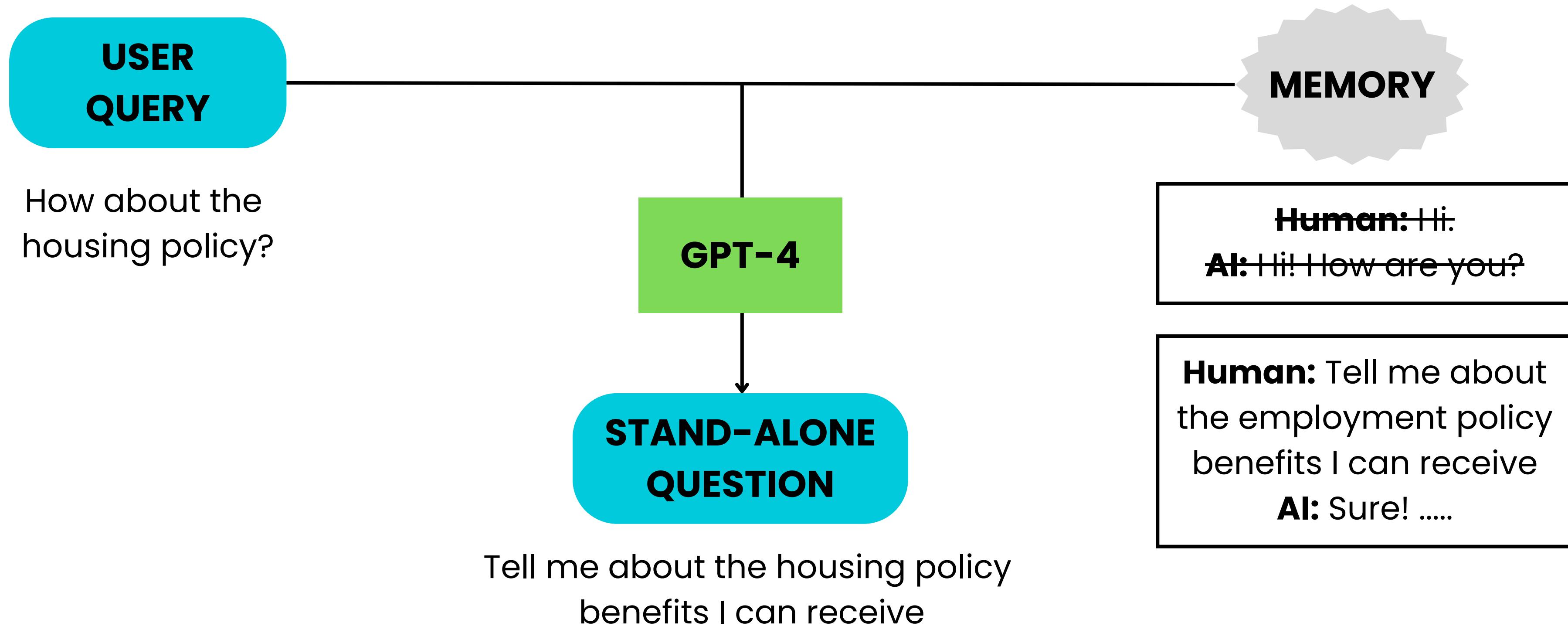
1. QUERY INPUT

**USER
QUERY**

How about the
housing policy?

MODEL PIPELINE

2. MAKE STAND-ALONE QUESTION



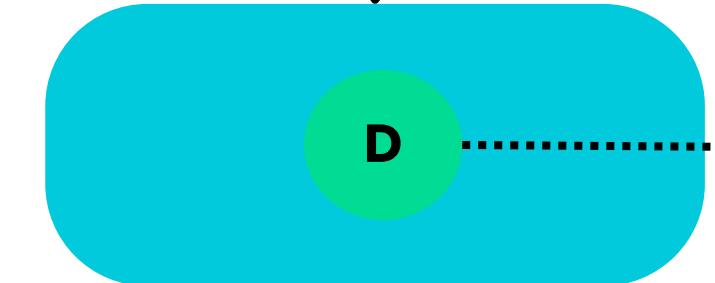
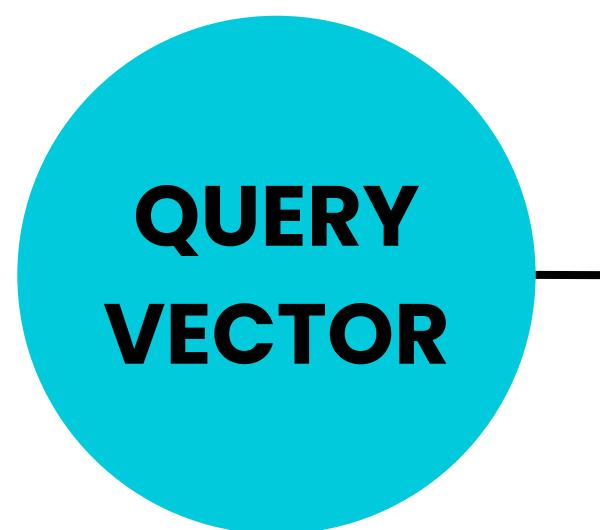
MODEL PIPELINE

3. MAKE QUERY VECTOR



MODEL PIPELINE

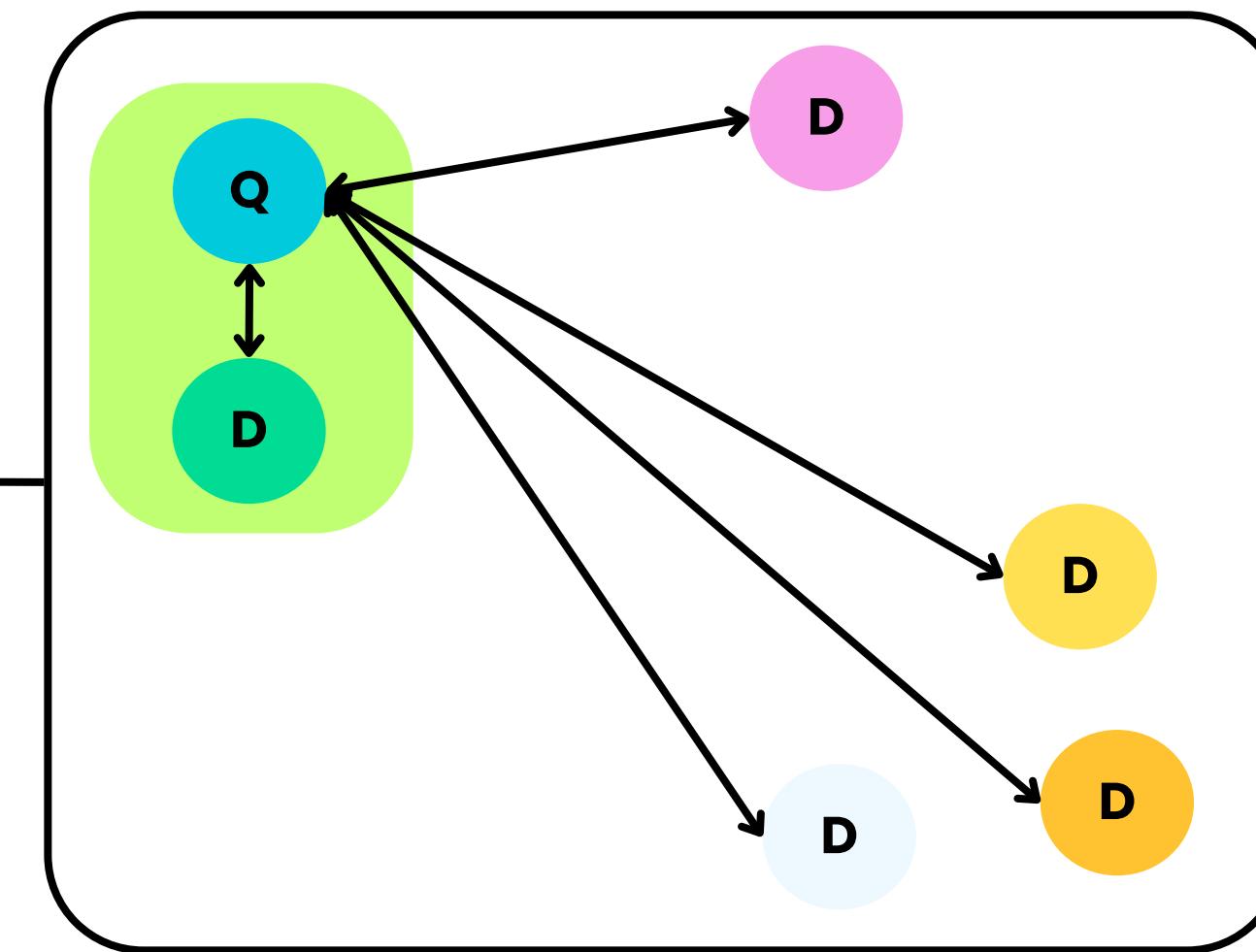
4. EMBEDDING-BASED RETRIEVAL



**RETRIEVED
DOCUMENTS**

($k = 1$)

VECTOR DATABASE



The content of the Suwon Happiness
Housing Policy is as follows:...

MODEL PIPELINE

5.CHAT COMPLETION

**STAND-ALONE
QUESTION**

Tell me about the housing policy
benefits I can receive

+

**RETRIEVED
DOCUMENTS**

The content of the Suwon Happiness
Housing Policy is as follows:...

GPT-4

RESPONSE

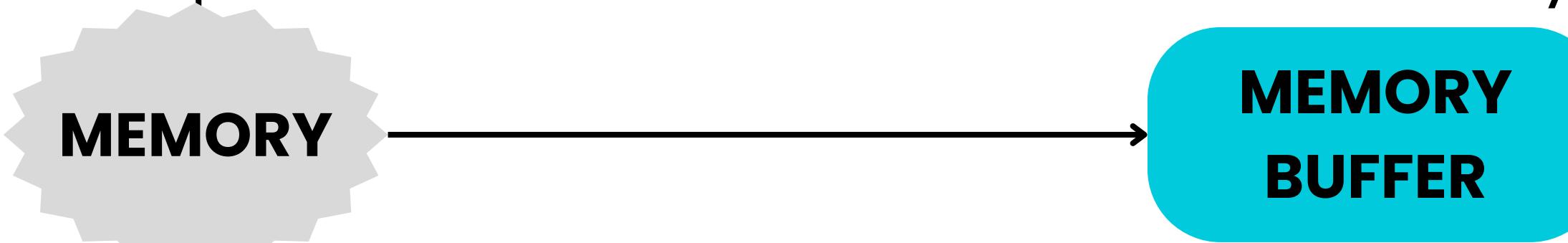
APPENDIX

PROMPT EXAMPLE – STAND ALONE QUESTION

Given the following conversation and a follow up question,
rephrase the follow up question to be a standalone question.

If you do not know the answer reply with 'I am sorry'.

Chat History:



Follow Up Input:



APPENDIX

PROMPT EXAMPLE – CHAT COMPLETION

Answer the user's questions using the given context.
If you do not know the answer reply with 'I am sorry'.

Context:

**RETRIEVED
DOCUMENTS**

Question:

**STAND-ALONE
QUESTION**